THE PROTEIN SOCIETY

Binding cavities and druggability of intrinsically disordered proteins

Yugang Zhang,[1,2] Huaiqing Cao,[1,2] and Zhirong Liu[1,2,3]*

[1]College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China
[2]Center for Quantitative Biology, Peking University, Beijing 100871, China
[3]Beijing National Laboratory for Molecular Sciences (BNLMS), Peking University, Beijing 100871, China

Received 31 October 2014; Accepted 12 January 2015
DOI: 10.1002/pro.2641
Published online 21 January 2015 proteinscience.org

Abstract: To assess the potential of intrinsically disordered proteins (IDPs) as drug design targets, we have analyzed the ligand-binding cavities of two datasets of IDPs (containing 37 and 16 entries, respectively) and compared their properties with those of conventional ordered (folded) proteins. IDPs were predicted to possess more binding cavity than ordered proteins at similar length, supporting the proposed advantage of IDPs economizing genome and protein resources. The cavity number has a wide distribution within each conformation ensemble for IDPs. The geometries of the cavities of IDPs differ from the cavities of ordered proteins, for example, the cavities of IDPs have larger surface areas and volumes, and are more likely to be composed of a single segment. The druggability of the cavities was examined, and the average druggable probability is estimated to be 9% for IDPs, which is almost twice that for ordered proteins (5%). Some IDPs with druggable cavities that are associated with diseases are listed. The optimism versus obstacles for drug design for IDPs is also briefly discussed.

Keywords: drug target; ligandability; druggability; drug design; intrinsically disordered protein; pE-DB; molecular recognition

Introduction

Although the study of intrinsically disordered proteins (IDPs) has a short history,[1–4] it was immediately recognized that such proteins are likely to be important targets in drug design.[5–7] IDPs are widely involved in critical cellular processes, including signal transduction and regulation[8] and are also associated with various human diseases.[9–11] Examples include the tumor suppressor p53, the breast cancer-related protein BRCA-1/2, the transcription factor c-Myc that is expressed constitutively in many cancer cells, α-synuclein that is related to various neurodegenerative diseases, and the tau protein in Alzheimer's disease. Statistically, 79% of cancer-associated proteins and 57% of the identified cardiovascular disease-associated proteins are predicted to contain disordered regions that are longer than 30 residues in length.[5,12] Consequently, IDPs are recognized as potential drug targets and are expected to play an active role in drug design.[6,7,13–16] However, compared with the well-developed drug design pipelines that target ordered (folded) proteins,[17] the drug designs that target IDPs remain in their infancy.[18] The studied IDP-related systems in drug design are limited and only a few small molecules and short peptides have been achieved to inhibit the function of IDPs.[19–24]

Therapeutic ligands usually accomplish their mission by binding to small cavities (binding sites or pockets) of target proteins. Before conducting drug

Abbreviations: CAVITY, the computer program to detect the binding cavities of proteins and quantitatively calculate their ligandability/druggability that was developed by Yuan, Pei and Lai; DisProt, database of protein disorder; IDPs, intrinsically disordered proteins; PDB, Protein Data Bank.

Grant sponsor: Ministry of Science and Technology of China; Grant number: 2015CB910300; Grant sponsor: National Natural Science Foundation of China; Grant number: 20973016.

*Correspondence to: Zhirong Liu, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China. E-mail: LiuZhiRong@pku.edu.cn

Table I. *List of the Dataset Disprot-pdb and the Determined Properties*

Disport ID	PDB & chain ID	Name	Complex	Method	Chain length	Conf. number	Disorder percent	Surface area (10^3 Å^2)	Cavity number	Druggable cavity probability
DP00040	2EZD.A	High mobility group protein HMG-I/HMG-Y	Hetero	NMR	21	1	100	3.03	0	0
DP00129	1IKN.A	Transcription factor p65	Hetero	XRD	285	1	100	16.1	10	0
DP00175	1JPW.D	Transcription factor 7-like 2	Hetero	XRD	38	1	100	2.97	0	0
DP00080	1KDX.B	Cyclic AMP-responsive element-binding protein 1	Hetero	NMR	28	17	100	2.92	0.71	0
DP00617	1MIU.B	26S proteasome complex subunit DSS1	Hetero	XRD	57	1	100	5.24	2	0
DP00213	1PJN.A	Histone-binding protein N1/N2	Hetero	XRD	21	1	100	2.96	0	0
DP00701	1RP3.B	Anti sigma factor FlgM	Hetero	XRD	87	1	95	6.36	3	0
DP00218	1S70.B	Protein phosphatase 1 regulatory subunit 12A	Hetero	XRD	291	1	100	17.9	7	0
DP00081	1TBA.A	Transcription initiation factor TFIID subunit 1	Hetero	NMR	67	25	100	5.43	2.72	0
DP00563	2BZW.A	Bcl2 antagonist of cell death	Hetero	XRD	196	1	100	7.56	5	0.2
n.a.	2LM0.A	Protein AF9 chimera	Hetero	NMR	79	10	100	11.4	5.5	0.091
DP00605	3FM7.C	Dynein intermediate chain, cytosolic	Hetero	XRD	27	1	100	3.73	0	0
DP00702	3KYS.B	Yes-associated protein (YAP)	Hetero	XRD	51	1	80	5.49	2	0.5
DP00365	1A17.A	Serine/threonine protein phosphatase 5	Homo	XRD	159	1	87	10.7	4	0
DP00132	1A8Y.A	Calsequestrin-1	Homo	XRD	345	1	100	18.7	11	0.091
DP00626	1AY9.A	Protein umuD	Homo	XRD	108	1	100	13.1	4	0.5
DP00637	1ET1.A	Parathyroid hormone	Homo	XRD	34	1	100	6.32	3	0.33
DP00723	2ZHI.A	Gibberellin receptor GID1A	Homo	XRD	315	1	100	15.1	7	0
DP00747_C002	1ANP	Atrial natriuretic factor	Mono	NMR	28	11	100	1.76	0.55	0
DP00588_C002	1CWX	Core protein p19	Mono	NMR	44	4	100	5.17	2	0
DP00071	1FTT	Homeobox protein Nkx-2.1	Mono	NMR	68	20	74.2	5.81	2.55	0
DP00729	1G6X	Pancreatic trypsin inhibitor	Mono	XRD	58	1	59	4.17	0	0
DP00335	1HN3	Cyclin-dependent kinase inhibitor 2A	Mono	NMR	40	20	100	5.17	2	0.15
DP00716_C001	1IVT	Lamin A/C	Mono	NMR	122	15	58	6.93	3.93	0
DP00730	1RRO	Oncomodulin	Mono	XRD	108	1	61	10.3	4	0
DP00423	1USS	Histone H1	Mono	NMR	88	10	93	6.43	4	0
DP00201	1VZS	ATP synthase-coupling factor 6, mitochondrial	Mono	NMR	76	34	100	6.61	4	0.014
DP00720	1WXL	FACT complex subunit Ssrp1	Mono	NMR	73	30	100	5.76	2.17	0
DP00549	1ZR9	Zinc finger protein 593	Mono	NMR	67	20	52	6.42	2.3	0.087
DP00717	1ZYI	Methylosome subunit pICln	Mono	NMR	116	15	58	9.06	4.93	0.30

Table I. *Continued*

Disport ID	PDB & chain ID	Name	Complex	Method	Chain length	Conf. number	Disorder percent	Surface area (10^3 Å^2)	Cavity number	Druggable cavity probability
DP00359	2DDN	Calvin cycle protein CP12	Mono	MD	80	1	100	6.01	2	0
DP00748_A002	2EYY	Adapter molecule crk	Mono	NMR	204	1	100	14.6	11	0.091
DP00748	2EYZ	Adapter molecule crk	Mono	NMR	304	1	51	18.2	13	0.15
DP00714	2JYP	Aragonite protein AP7	Mono	NMR	36	1	100	2.86	1	0
DP00646	2K7M	Gap junction alpha-5 protein	Mono	NMR	109	10	100	12.3	4.5	0.022
DP00622	2KOG	Vesicle-associated membrane protein 2	Mono	NMR	116	20	79	13.1	3.4	0.15
DP00534	2LJ9	Calvin cycle protein CP12-2, chloroplastic	Mono	NMR	22	20	100	2.38	0.5	0

Surface area and cavity number are averaged on conformations for each protein. Surface area is calculated using the PyMOL package as the solvent-assessable surface area with a probe radius of 1.4 Å.

design on a particular protein, it is important to assess its possibility to be a good target, for example, whether the protein has suitable geometrical shaped cavities for ligand binding. This is known as the "druggability" or "ligandability" assessment problem in drug discovery.[25] By testing a few datasets, Yuan *et al*. found that developed detection methods are not only able to detect the binding cavities on the protein surface, but can also discriminate druggable cavities from less druggable ones based on protein structures with considerable accuracy.[25] For IDPs, although they do not have ordered structures in the free state under physiological conditions, they may undergo a disorder-to-order transition upon binding to their biological partners via coupled folding and binding.[26] Analyses on the solved structures of IDPs in complexes can give valuable molecular interaction information of IDPs.[27,28] For example, it was shown that IDPs possess greater surface and interface areas per residue than ordered proteins,[28] and the interface structure of IDPs is more dynamic than that of ordered proteins.[29] Considering the structural difference between IDPs and ordered proteins, it would be intriguing to investigate whether IDPs afford binding cavities and druggability different from well folded ordered proteins.

In this article, we conducted a comparative study on the binding cavities and druggability of IDPs and ordered proteins. IDPs are predicted to possess more binding cavities than ordered proteins of a similar length, and their cavity geometries are different. Most importantly, the druggability of the cavities of IDPs may be comparable with those of ordered proteins, which sheds optimistic light on the drug design toward IDPs.

Results

Data for analysis

Three datasets were used in our analysis: Disprot-pdb and pE-DB for IDPs, and CavityTEST for ordered proteins. We constructed Disprot-pdb by scanning the DisProt and PDB to select proteins with at least 50% of solved amino acids in the PDB structure being shown disordered in DisProt. The Leukemia fusion target AF9, against which inhibitors have been designed, was also included in the Disprot-pdb. pE-DB was adopted from Varadi *et al*., which provided structural ensembles of some IDPs.[30] In comparison with those in Disprot-pdb, the structures in pE-DB are less accurate, but the average conformation number for one protein is much larger in the latter and this greatly facilitates the analysis of property distribution for a protein. We also included the oncoprotein c-Myc into the pE-DB, whose binding sites for ligands have been identified in experiment[21] and the conformational ensemble has been characterized by large-scale molecular dynamics simulations.[18]

Table II. *Information of the Dataset pE-DB*

pE-DB ID	Name	Method	Chain length	Conf. number	Surface area (10^3 Å^2)	Cavity number	Druggable cavity probability
1AAA	Phosphorylated Sic1	SAXS & NMR	92	32	10.5	3.5	0.071
1AAB	Heat shock protein beta-6 (HSPB6) fragment (57–160) V67G mutant	SAXS	208	5	14.5	7.2	0
2AAA	Unbound $p27^{KID}$ domain	MD	69	130	7.06	4.03	0.011
2AAB	Heat shock protein beta-6 (HSPB6) fragment (24–160)	SAXS	274	8	19.8	9.38	0.093
3AAA	CYNEX4 flexible multidomain FRET probe	SAXS	825	17	37.9	24.18	0.027
3AAB	Heat shock protein beta-6 (HSPB6) fragment (40–160)	SAXS	484	4	30.6	18.5	0.20
4AAA	CYNEX4 T266 mutant flexible multidomain FRET probe	SAXS	825	16	38.2	24.13	0.034
5AAA	ParE2-associated antitoxin (PaaA2)	SAXS & NMR	71	50	8.31	2.52	0.16
5AAC	Phosphorylated Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase	SAXS & NMR	666	44	38.7	19.91	0.081
6AAA	$p15^{PAF}$	SAXS & NMR	110	4939	12.6	4.90	0.081
6AAC	K18 domain of Tau protein	NMR	130	995	11.5	3.22	0.0028
7AAA	Heat shock protein beta-6 (HSPB6)	SAXS	320	6	25.7	10.17	0.30
7AAC	N-TAIL Measles nucleoprotein	NMR	132	995	11.4	3.27	0.014
8AAA	Heat shock protein beta-6 (HSPB6) fragment (57–160)	SAXS	416	3	25.5	17.33	0.15
9AAA	Sic1	SAXS & NMR	92	44	9.31	3.68	0
n.a.	c-Myc	MD	40	308	3.77	1.43	0.011

Disprot-pdb and pE-DB contained 37 and 16 entries, respectively. CavityTEST is a dataset adopted from Laurie *et al.*,[31] containing 35 structurally distinct ordered proteins that have been determined to bind ligands. The datasets are listed in Tables I to III. It is noted that the three datasets are markedly different: CavityTEST contains stable protein structures; nearly half the Disprot-PDB entries (18 out of 37) describes IDPs bound to their partners (in a structural state which does not necessarily exist in solution), while pE-DB contains structural ensembles of IDPs in solution which are not unique. It is remarkable that the cavity properties of the three distinct datasets come out to be comparable as described in the follows.

The protein structures in the datasets were analyzed using the program CAVITY developed by Yuan *et al.*[25] to give information on their binding cavities, for example, the cavity number, their geometries and the druggability. A brief description on CAVITY is provided in Materials and Methods.

Cavity number

The cavity numbers for all examined protein are listed in Tables I to III. It is noted that these numbers are averages for many distinct conformations when conformation ensembles are available in Disprot-pdb and pE-DB. For the ordered protein dataset CavityTEST, each protein has at least one predicted cavity, being consistent with its collection criterion that the proteins were experimentally determined to bind ligands. For Disprot-pdb, only five proteins were predicted to have no binding cavity. Interestingly, the structures of all these five proteins were solved via X-ray diffraction (XRD), and four of them were in heterocomplexes.

At first glance, the average cavity number for proteins in Disprot-pdb (3.65) is much smaller than that in CavityTEST (8.37) or pE-DB (9.83), but this is misleading because of the shorter chain length of the protein sequences in Disprot-pdb. When we plot the cavity number as a function of the chain length (Fig. 1), it can be seen that the cavity number in Disprot-pdb is close to that of pE-DB and is slightly larger than that of CavityTEST under the same chain length [Fig. 1(a)]. We define a new quantity, the cavity number per 100 residues, to describe such a trend. The scattering data are plotted in Figure 1(b), which fluctuates dramatically at short lengths and converges at longer sequences. Numerically, the

Table III. *Information of the Dataset CavityTEST*

PDB ID	Chain length	Surface area (10^3 Å^2)	Cavity number
1A4J	653	41.0	19
1A6U	228	10.1	3
1AHC	246	11.4	7
1BBS	660	29.1	16
1BRQ	174	9.43	7
1BYA	491	19.5	15
1CGE	162	8.45	5
1CHG	226	9.78	5
1DJB	257	11.5	10
1HSI	198	10.6	9
1IFB	131	7.07	1
1IME	544	21.5	16
1KRN	89	4.88	2
1L3F	316	12.2	5
1NNA	387	14.9	12
1PDY	433	16.7	8
1PHC	405	18.5	15
1PSN	326	13.7	3
1PTS	234	10.8	8
1QIF	532	20.8	14
1STN	136	7.88	4
1YPI	494	19.8	19
2CBA	258	11.8	9
2CTB	307	12.1	8
2PTN	223	9.34	5
2RTA	131	7.59	3
2SIL	381	14.6	11
2TGA	223	9.29	4
3APP	323	14.9	3
3LCK	288	14.6	12
3P2P	238	13.3	11
4CA2	255	11.5	10
5CPA	307	12.4	8
6INS	100	6.35	2
7RAT	124	6.91	4

average cavity number per 100 residues of Disprot-pdb and pE-DB is 3.40 and 3.31, respectively, about 20% larger than the value for CavityTEST (2.80). This observation is likely to be due to IDPs affording greater surface area per residue than ordered proteins (Fig. 2), as already suggested.[27,28] Therefore, IDPs are predicted to possess more binding cavities than ordered proteins of a similar length. This observation supports the concept that IDPs economize genomes and protein resources,[32] that is, IDPs are capable of using smaller protein size to afford the same interface area as ordered proteins.

IDPs usually exist in an ensemble of rapidly changing conformations and exhibit almost unlimited structural heterogeneity.[33–35] The cavity properties will likely vary between the various conformations the protein adopted. For each conformation (structure), a cavity number can be determined with CAVITY. Therefore, a distribution of cavity number is got for any given protein with multiple PDB structures. In Figure 3, we have examined the distribution of the cavity number for the conformation ensemble of a few IDPs. The distribution is found to be wide in all cases. In particular, for proteins from pE-DB, where there are a significant number of conformations (up to a few hundred or thousand) available in the database, the distribution can be well described by a Gaussian function [solid lines in Fig. 3(d–f)]. Such heterogeneity may have essential effects on rational drug design of IDPs. On the one hand, heterogeneity of cavity number would be accompanied by heterogeneity of cavity shape, position on the chain, and chemical properties, which could aid rational drug design by increasing the number of distinct targets. On the other hand, the heterogeneity of cavity suggests that it is difficult to use a single ligand to bind all various conformations of an IDP, which is distinct from the case of ordered proteins. The binding should be considered in terms of conformation selection or induced fit. In addition, the heterogeneity of cavity for an IDP may result in lower specificity, which also hampers the rational drug design.

Cavity geometries

Owing to their high chain flexibility, IDPs have extended structures that are significantly less flat

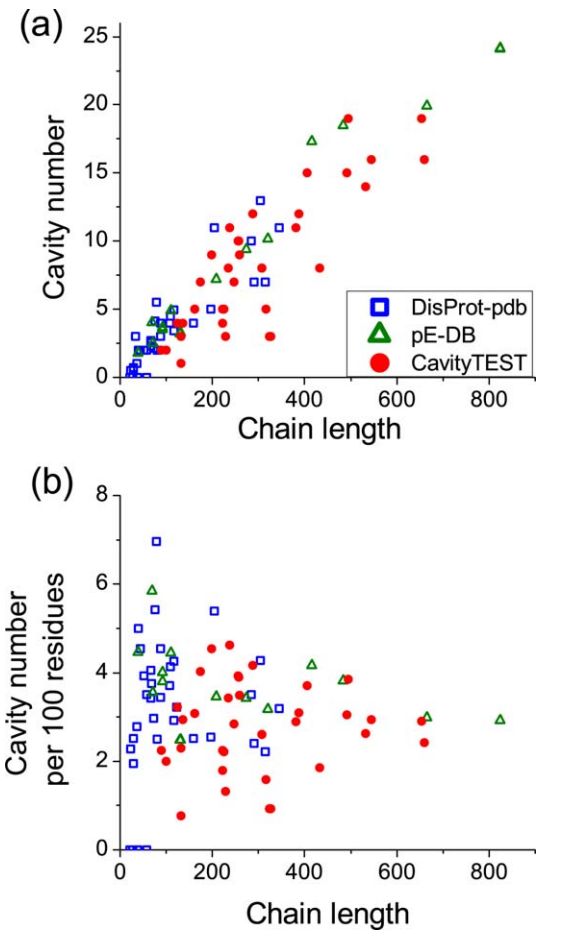

Figure 1. (a) Cavity number and (b) cavity number per 100 residues as a function of the protein chain length (residue number) for proteins in Disprot-pdb (blue squares), pE-DB (green triangles) and CavityTEST (red circles).

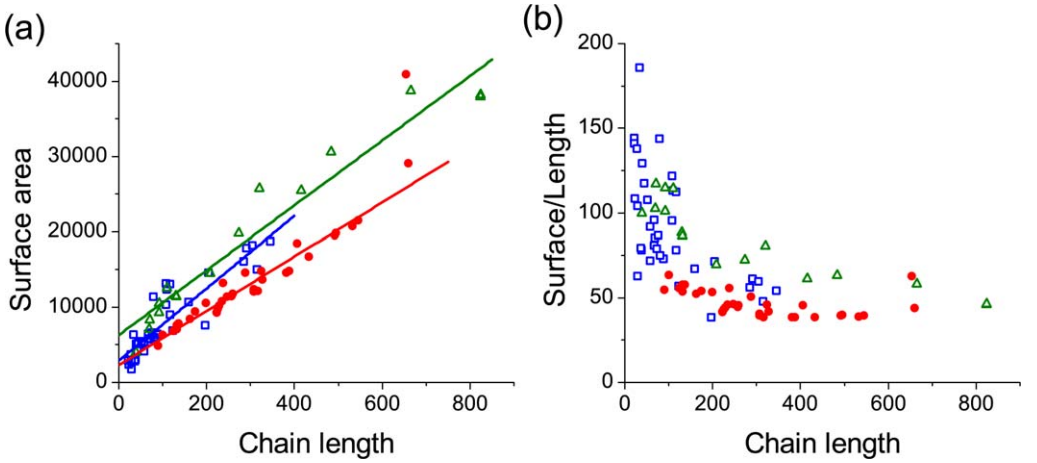

Figure 2. (a) Protein surface area (Å^2) and (b) the ratio between surface area and chain length (i.e., the surface area per residue, Å^2) as a function of the protein chain length (residue number) for proteins in Disprot-pdb (blue squares), pE-DB (green triangles), and CavityTEST (red circles).

than those of ordered proteins.[32] Thus, it is expected that IDPs have larger cavities than ordered proteins. This is validated in our analysis of the cavity surface area and volume (Fig. 4 and Table IV). The distributions of the surface area and volume exhibit a peak at low values and a long (fat) tail at high values [Fig. 4(a,b)]. In comparison with the datasets for IDPs (Disprot-pdb and pE-DB), the dataset for ordered proteins (CavityTEST) has a higher peak at the low value. As a result, the average cavity surface area (volume) of Disprot-pdb is 13% (28%) larger than that of CavityTEST, and the values for pE-DB are even larger (Table IV). We also calculated the depth of the cavities, which showed that the average cavity depth of IDPs in the Disprot-pdb is similar to those found for proteins in the CavityTEST dataset. In addition, the average cavity surface area and volume for proteins were found to obey the inherent scaling law of $S \propto V^{3/2}$ [Fig. 4(c)].

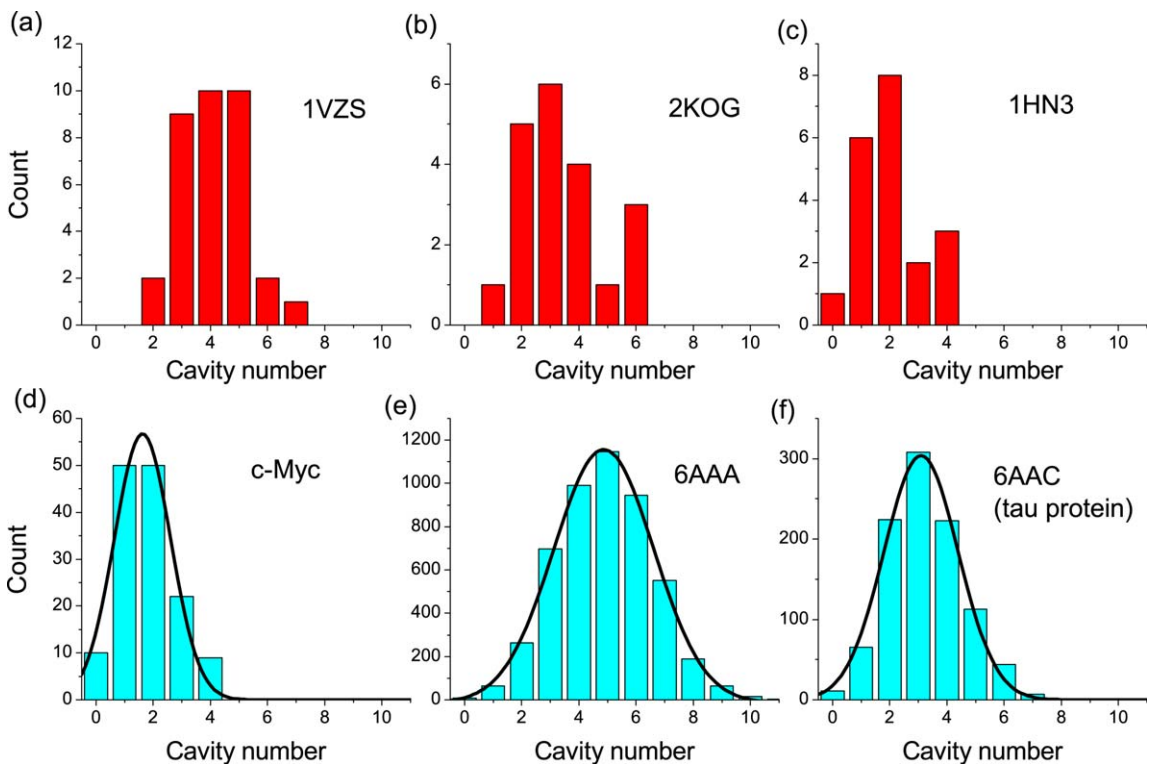

Figure 3. Distribution of the cavity number for the conformation ensemble of a few representative proteins. Systems in (a–c) belong to Disprot-pdb and those in (d–f) belong to pE-DB. Solid lines in (d–f) are fits to the scattering data with a Gaussian function.

Table IV. *Average Properties for Proteins in Three Datasets*

Dataset	Chain length	Cavity number	Cavity number per 100 residues	Cavity surface area (Å^2)	Cavity volume (Å^3)	Cavity depth (Å)	Cavity segment number	Cavity segment length	Druggable cavity probability	Cavity conservation p_{com}	Cavity RMSD (Å)
Disprot-pdb	107.2	3.65	3.40	338	513	4.4	4.3	7.0	0.091	0.57	2.8
pE-DB	297.1	9.83	3.31	462	788	5.1	4.7	7.7	0.092	0.52	3.5
CavityTEST	299.4	8.37	2.80	299	399	4.4	5.3	5.5	0.055	–	–

Another property of the cavity is the segment number, that is, how many continuous segments are assembled to constitute the cavity. The cavities of IDPs were found to be less fragmented, that is, they are assembled from fewer segments than those of ordered proteins (Fig. 5). On average, the segment number for cavities in Disprot-pdb is 4.3, which is 23% smaller than that of CavityTEST (5.3). Ordered proteins hardly ever use a single segment to constitute a cavity, and the occurrence of two segments is also much lower than that in IDPs. The origin for such a difference arises from the free energy paid in bringing distant segments into close proximity for IDPs owing to their inherent chain flexibility. Similar differences were also observed in protein-protein interaction interfaces for IDPs and ordered proteins by Meszaros *et al*.[28] The smaller segment number of IDPs is helpful for the economization of protein resources, that is, IDPs can use shorter sequence to create interface or cavity of similar size as that for ordered proteins.

Two examples of IDPs with less-fragmented cavities are given in Figure 6. The illustrated conformation of the cyclin-dependent kinase inhibitor 2A possesses one cavity with a volume of 543 Å^3. Vesicle-associated membrane protein 2 contains two cavities, with a volume of 1715 and 1645 Å^3. All three cavities are constituted by a single continuous segment, respectively.

The conservation of cavities is a critical issue in a conformation ensemble with many distinct conformations. The meaning of a cavity would be much stronger if it was present in many conformations. We measured the conservation of cavities in terms of the proportion of common atoms (p_{com}) and the RMSD values (see Materials and methods). The average p_{com} for Disprot-pdb and pE-DB are 57% and 52%, respectively. Such conservation is higher than expected, especially for pE-DB which possesses diverse conformations. On the other hand, the average cavity RMSD are 2.8 and 3.5 Å for Disprot-pdb and pE-DB, respectively.

Cavity druggability

Based on the geometrical structure and physical chemistry properties, CAVITY can give a predicted average binding pK_d of the binding cavity with properly designed ligands in general (see Materials and Methods for details).[25] If the predicted pK_d is less than 6.0, the cavity may not be a suitable drug design target.[25] The summary of the predicted pK_d values for three datasets are plotted in Figure 7. The fraction of cavities to have a predicted p$K_d > 6.0$ is about 30% for IDPs (both Disprot-pdb and pE-DB), whereas the value for ordered proteins (CavityTEST) is only 20%. On the other hand, the fraction with unfavorable p$K_d < 5.0$ for ordered proteins is as high as 40%, greatly exceeding that for IDPs (24% for Disprot-pdb

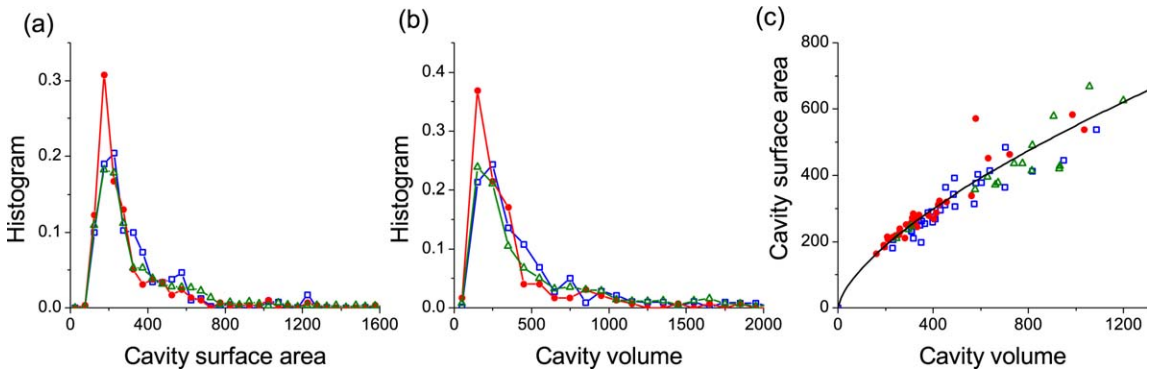

Figure 4. (a) Distribution of the surface area (in a unit of Å^2) of cavities. (b) Distribution of the volume (in a unit of Å^3) of cavities. (c) Correlation between the average cavity surface area and volume for proteins. Data are presented for Disprot-pdb (blue squares), pE-DB (green triangles) and CavityTEST (red circles). Solid line in (c) is a trend description of the scattering data with a form of $y = 5.5\ x^{2/3}$.

and 30% for pE-DB). The entropy effect due to coupled folding and binding was not considered in the pK_d prediction by CAVITY, which may decrease the actual pK_d of IDPs because the binding effect of a small molecule would be partially compensated by the conformational adjustment of IDPs.[36] Nonetheless, the results here reflect an optimistic possibility of designing small active ligands that interact with IDPs.

Affinity is a necessary but not sufficient condition for druggability since protein druggablity is a more complicated property that is affected by various factors at a system biological level. An algorithm has been developed in CAVITY to classify the druggability of cavities into three types (druggable, amphibious and undruggable) with considerable prediction accuracy.[25] We determined the druggability type of each cavity. The druggable cavity probability for each protein of the Disprot-pdb and pE-DB datasets is listed in Tables I and II. The determined druggable probability lies between 0 and 0.5, and

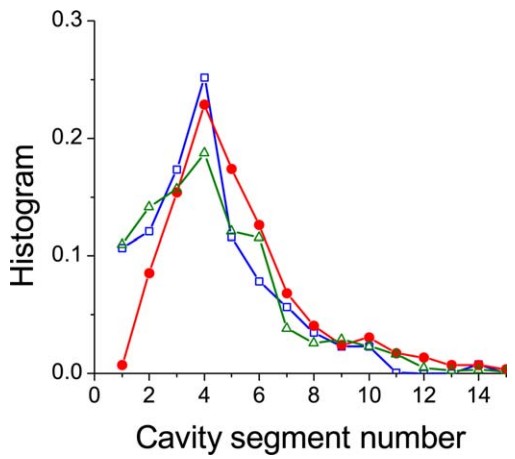

Figure 5. Distribution of the number of non-continuous sequence segments of the cavity, given for Disprot-pdb (blue squares), pE-DB (green triangles), and CavityTEST (red circles).

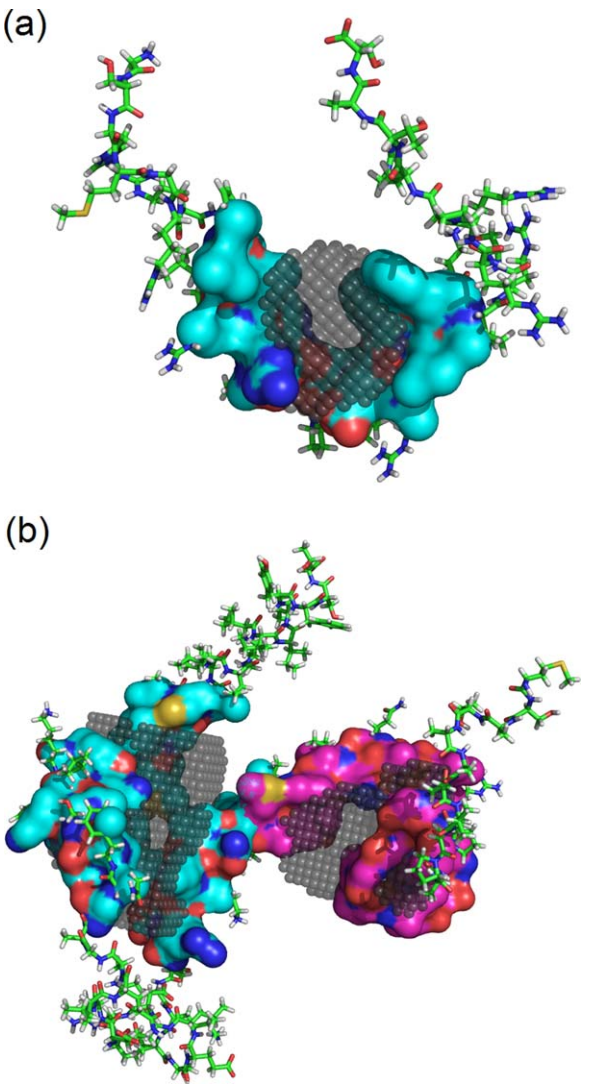

Figure 6. Two examples of IDPs where the cavities are composed of a single segment. (a) Cyclin-dependent kinase inhibitor 2A (PDB ID 1HN3) is a tumor suppressor. (b) Vesicle-associated membrane protein 2 (PDB ID 2KOG). The chain segments constituting the cavities are shown in color-spheres and surface, and the cavity vacant is filled with gray small balls. Graphics is prepared using PyMOL.

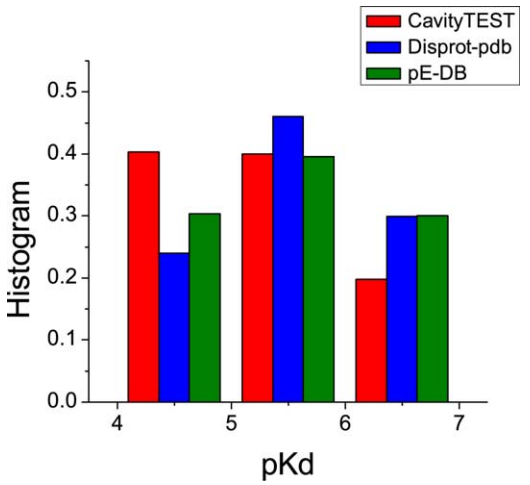

Figure 7. Distribution of the predicted binding pK_d for cavities in three datasets.

the highest value (0.5) is achieved in the Yorkie homolog (PDB ID 3KYS) and umuD (PDB ID 1AY9). Overall, the average druggable probability is 9% for proteins in both Disprot-pdb and pE-DB datasets, which is essentially double the value observed for the CavityTEST dataset (5%; see Table IV). This difference is even more pronounced than the pK_d values discussed above. Molecular recognition functions of IDPs are closely related to their molecular recognition features (MoRFs).[37,38] Using a predictor named MoRFpred,[39] we have checked the predicted MoRFs of a few systems in the datasets and found some correlation between the predicted MoRFs and druggable cavities (data not shown).

Discussion

Cavities in the rapidly fluctuating ensemble

The above analyses on IDPs were conducted in a manner similar to those on ordered proteins. Since IDPs exist in highly dynamic conformations, some concerns may be raised, which should be appropriately addressed.

One point is the suitability to define a cavity in the rapidly fluctuating ensemble of conformations that an IDP samples. The energy landscapes of IDPs are relatively flat and the conformations interconvert very fast, for example, in a timescale of nanosecond, which is much faster than the typical binding time of a ligand. Therefore, is it meaningful to define a cavity in IDPs? Would a predicted cavity wither away far before a ligand succeeds to bind it? The answer to dismiss such concern roots in the statistical thermodynamics: equilibrium population is governed by such laws as Boltzmann distribution and does not depend on the kinetic process. If a single conformation of protein can bind a ligand, the identical conformation in an ensemble can do the same thing although the conformation is now accompanied by a weight determined by the ensemble. Different kinetic schemes are possible in affording the thermodynamics. For example, after the ligand bind a short-lived conformation with an appropriate cavity, it may lock the protein in such a conformation, or, force the protein to jump among conformations with a similar cavity, but not to those with improper cavities. No matter how the kinetics comes out to be, the thermodynamics does not alter.

Another one is the suitability of the dataset of IDPs used. Some structures (although not all) in Disprot-pdb came from complexes by removing the partners. So one might question whether they can reflect the properties of IDPs in the disordered free form which would be the target. Here, we note that structures in pE-DB are mostly in the disordered free form, and the resulting difference with respective to ordered proteins is in the same direction with Disprot-pdb. Therefore, although the obtained quantitative values can not be considered accurate, the qualitative conclusions are likely reliable.

Examples of drug design of IDPs

IDPs are abundant in cells, but drug design where IDPs are the target remains an untapped source. Here we briefly survey examples of IDP drug design in the literature (Table V) and discuss their druggability when data are available.

There are a few examples that are widely discussed in reviews,[6,7,15,47–59] namely, p53-MDM2, c-Myc-Max, and EWS-Fli1.[20,40,45] The tumor-suppressor protein p53 is at the center of a large signaling network involved in cell cycle control, senescence, and apoptosis in response to oncogenic or other cellular stress signals.[60,61] The p53 protein is regulated by binding with multiple targets such as MDM2 and Taz2.[62] Small molecules have been screened to inhibit p53-MDM2 interaction and reactivate the p53 pathway in cancer cells.[40–42] These small molecules function by binding to MDM2 in the p53-binding pocket, but do not interact directly with p53. Therefore, this example belongs to "drug design involving IDPs," but not "drug design targeting IDPs." EWS-Fli1 and c-Myc-Max, on the other hand, belong to the latter case. EWS-Fli1 is an oncogenic fusion protein, which is exclusively present in Ewing's sarcoma family tumors.[20] C-Myc is a transcription factor that becomes active by forming a dimer with its partner protein Max, and is expressed constitutively in most cancer cells.[21] Both c-Myc and EWS-Fli1 are IDPs. By systematic screenings, small molecule inhibitors were identified that bind to c-Myc and EWS-Fli1 directly and prevent their interaction with partners.[20,21,43–48] The conformation ensemble of c-Myc$_{370–409}$ in the unbound state has been characterized by MD simulations,[18] which was included in our pE-DB dataset. Five conformations with druggable cavities were identified in our

Table V. *Summary of IDP-Targeting Drug Development Efforts*

System	Therapeutic importance	Inhibitor	Target against the IDP or its partner?	Efforts	Reference
p53-MDM2	p53 is the most commonly mutated gene in human cancer. Pharmacologic activation of the p53 pathway is highly important for therapeutics.	Nutlin	Ordered partner	IC_{50} is 100–300 *nM*. When used at μ*M* concentrations, nutlin arrested proliferating cancer cells and induced apoptosis in a number of different cancer cell lines including colorectal, lung, breast, prostate, melanoma, osteosarcoma and renal cancer. It is currently in phase-I clinical trials.	40–42
c-Myc-Max	c-Myc is a seldom-mutated transcription factor. Its deregulated expression is associated with numerous types of human cancers.	Peptidomimetic inhibitors	IDP	IC_{50} is about 50 μ*M* in the ELISA and EMSA assays. Inhibit the cell foci formation in cultures of chicken embryo fibroblasts (CEF) with $IC_{95} = 20$ μ*M*.	43,44
		10058-F4 and 10074-G5	IDP	Binding Kd is 2–20 μ*M*. IC_{50} in 5–50 μ*M* in a cell-based proliferation assay with HL60 human promyelocytic leukemia cells which overexpress Myc due to gene amplification. When administered at 20 mg/kg (∼300 μ*M*) as a daily dose, it can double the survival of mice genetically engineered to develop neuroblastoma.	21,45–47
EWS-Fli1	EWS-Fli1 is an oncogenic fusion protein, which is exclusively present in Ewing's sarcoma family tumors.	YK-4–279	IDP	Binding Kd is ∼10 μ*M*. IC_{50} to the growth of EWS-FLI1–positive Ewing's sarcoma family tumors (ESFTs) cell line is 900 n*M*. A 72 mg/kg of daily dose (maintaining 3 μ*M* level) significantly deduces the tumor size in rat.	20,48
AF9-AF4	AF9 is a mixed lineage leukemia (MLL) fusion protein that causes oncogenic transformation of hematopoietic cells. It interacts with AF4, the most common fusion protein in acute leukemias. AF4–AF9 protein complex is a promising target for leukemia therapy.	Peptide PFWT	IDP	At 10 μg/mL, PFWT completely blocks AF4–AF9 binding *in vitro*. At 25 μg/mL, it induces death by necrosis in the t(4;11) leukemia cell line. Treatment of the cell line with PFWT in combination with four chemotherapeutic compounds results in sequence-dependent synergy, suggesting that PFWT can augments the effects of several clinically available chemotherapeutic agents for MLL leukemias.	22,49,50
		Non-peptide compounds	IDPs	18 compounds were identified in a competitive screening assay, with IC_{50} of 3–50 μ*M*.	51
PTP1B	PTP1B is a negative regulator of insulin and leptin signaling. It is a validated therapeutic target for diabetes, obesity, and breast cancer.	MSI-1436	Disordered terminus	Inhibits the enzyme function of PTP1B with $K_i = 600$ n*M*. In a xenograft model, the mice with a dose of 5 mg/kg every 3 days displayed a marked decrease in tumor size and tumor number.	52

Table V. *Continued*

System	Therapeutic importance	Inhibitor	Target against the IDP or its partner?	Efforts	Reference
Aggregating IDPs	Pathological aggregation of IDPs triggers a series of human neurodegenerative diseases, for example, Alzheimer's disease, Down's syndrome, Parkinson's disease and prion diseases.	Molecular tweezers	IDPs	Molecular tweezers bind to lysine- and arginine-containing small peptides with K_d = 20–100 μ*M*. They inhibit the aggregation and toxicity of multiple amyloidogenic proteins in cell culture with IC_{50} of 3–50 μ*M*. In a novel zebrafish model of α-syn toxicity, 10 μ*M* improved survival by threefold at 72 h postfertilization (hpf) and 13-fold at 240 hpf.	53–55
		Non-natural amino-acid peptide D-TLKIVW	Amyloid form of IDPs	The apparent K_d to tau fibrils is ~2 μ*M*. In tenfold molar excess, it prevents the fibril formation in the presence of preformed fibril seeds.	23
		ELN484228	IDPs	It did not detectably modify αSyn aggregation *in vitro*, but it is protective in cellular models, for example, to restore phagocytosis in the αSyn overexpressing cells by 60% with a 30 μ*M* solution.	56

analysis for c-Myc (Fig. 8). Hammoudeh *et al*. have experimentally identified the binding sites of different inhibitors in c-Myc.[21] These actual binding sites correlate well with the druggable cavities predicted by CAVITY (Fig. 8).

AF9 is a mixed lineage leukemia (MLL) fusion protein that causes oncogenic transformation of hematopoietic cells.[36] AF9 interacts with AF4, the most common fusion protein in acute leukemias. Bioinformatics analysis has revealed that fusion proteins are usually significantly enriched in structural disorder.[64] In the current example, both AF9 and AF4 are IDPs. Based on mapping studies, an AF4-derived peptide has been developed to specifically interact with AF9 and disrupt the AF4-AF9 interaction *in vitro* and *in vivo*.[22] The peptide induces necrotic cell death in leukemia cells

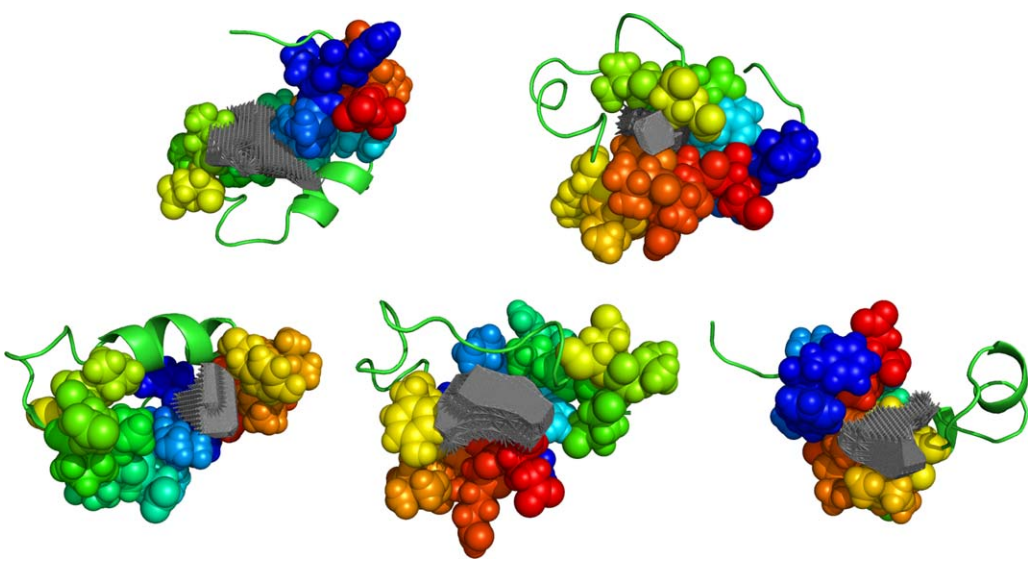

Figure 8. Five conformations of c-$Myc_{370–409}$ with druggable cavity. Experimentally suggested binding sites (residues 374–385, and 402–409) are shown in spheres with rainbow colors, while other residues are shown in cartoon. The bottoms of the druggable cavity are depicted in dense gray lines. Graphics is prepared using PyMOL.

Table VI. *List of IDPs as Potential Drug Targets*

PDB/pE-DB ID	Short name	Biological function and relation to diseases	Reference
2BZW	Bcl2	The phosphorylated form is anti-apoptotic, and the dephosphorylated form is pro-apoptotic. The latter may be involved in neural diseases such as schizophrenia.	69
2LM0	AF9	A mixed lineage leukemia (MLL) fusion protein that causes oncogenic transformation of hematopoietic cells.	63
3KYS	YAP	Transcriptional coactivator, a key regulator of organ size and a candidate human oncogene inhibited by the Hippo tumor suppressor pathway.	70
1A8Y	Calsequestrin	The major Ca^{2+} storage protein of muscle. Has significant affinity for a number of pharmaceutical drugs with known muscular toxicities.	71
1ET1	Parathyroid hormone	A hormone to increase the concentration of Ca^{2+} in the blood. In excessive amounts it is the main character of hyperparathyroidism.	72
1HN3	$p19^{Arf}$	Promotes p53-mediated cell arrest and apoptosis. It is a frequent target for loss in human cancers.	73
1ZR9	ZNF593	Negative modulator of the DNA binding activity of the Oct-2 transcription factor. Associated with malaria and prostatitis.	74
1ZYI	pICln	A multifunctional protein involved in regulatory mechanisms as different as membrane ion transport and RNA splicing.	75
2EYY/2EYZ	crk	Regulates transcription and cytoskeletal reorganization during cell growth, motility, proliferation and apoptosis. Responsible for the malignant features of various human cancers.	76
2KOG	synaptobrevin	Participates in neurotransmitter release at a step between docking and fusion. Likely related to familial infantile myasthenia.	77
2AAA	$p27^{KID}$	Influences cell division by regulating nuclear cyclin-dependent kinases. Its reduced levels in most cancers are correlated with increased tumor size and increased tumor grade.	78
6AAA	$p15^{PAF}$	A proliferating cell nuclear antigen (PCNA)-associated protein overexpressed in multiple types of human cancer.	79
6AAC	tau	Stabilize microtubules. Abundant in neurons of the central nervous system. Responsible for Alzheimer's disease.	80
n.a.	c-Myc	A transcription factor that activates expression of many genes. Persistently expressed in many cancers.	18

and enhances the cytotoxic effect of established chemotherapeutic agents, holding promise as a component in the composite therapy for MLL leukemia.[49,50] Recently, nonpeptidic inhibitors of AF9 were also successfully developed by a high-throughput screening assay.[51] AF9 was included in our Disprot-pdb dataset (PDB ID 2LM0). The 10 conformations of AF9 in the PDB afforded 55 cavities; 5 of them were predicted to be druggable (Table I). Therefore, AF9 is highly druggable.

PTP1B (protein-tyrosine phosphatase 1B) is a nontransmembrane enzyme found on the endoplasmic reticulum. It is a negative regulator of insulin and leptin signaling. PTP1B has been long recognized as a therapeutic target for diabetes and obesity.[65] In addition, it is overexpressed in breast tumors together with HER2, and its overexpression alone drives mammary tumorigenesis. Therefore, PTP1B acts also as a therapeutic target for mammary tumorigenesis and malignancy. PTP1B

contains an ordered catalytic domain and a long disordered C terminus. Recently, an aminosterol natural product, trodusquemine (MSI-1436) was found to inhibit the enzyme function of PTP1B by binding to its disordered C terminus.[52] Interestingly, MSI-1436 works via an allosteric effect, that is, it binds two sites that are distinct from the active enzyme site and stabilizes an inactive conformation of PTP1B. This is in accordance with the suggestion that allostery has direct implications for the role of structural disorder in proteins and is thus helpful for the development of drugs and therapies.[66]

Some other progress was achieved in targeting aggregating IDPs.[23,59] Although the majority of IDPs have an inherent advantage in preventing aggregation,[32,56] some "abnormal" IDPs are commonly found among proteins involved in amyloid formation and conformational diseases. The suppression of pathological amyloid fibril formation is an active area of research, and some strategies have been explored. For example, molecular tweezers were found to effectively perturb the aggregation processes via specific binding to lysine.[53–55] The known atomic structures of segments of amyloid fibrils were also used as templates in designing non-natural amino-acid inhibitors of amyloid fibril formation,[23] and a virtual screening was conducted on a subset of αSyn conformations to identify a ligand that is protective in cellular models of αSyn-mediated vesicular dysfunction.[56,68]

IDPs as potent drug targets

To be a potent drug target, the protein should not only have the potential to interact with designed small ligands, but should also possess an essential biological function and be closely related to diseases. Based on the druggable cavity probability of the proteins as discussed above and their biological importance in the literature, we provide a list of IDPs in Table VI that are suitable targets for rational drug design. A few systems are discussed briefly as follows.

Adapter molecule crk (PDB ID 2EYY/2EYZ) is also known as proto-oncogene c-Crk or p38. It has several SH2 and SH3 domains and acts as an adaptor to link tyrosine kinases and small G proteins. It regulates transcription and cytoskeletal reorganization during cell growth, motility, proliferation, and apoptosis.[76] Increased expression of crk has been identified to be responsible for the malignant features of several human cancers including breast, ovarian, lung, brain, and stomach. Therefore, the inhibition of crk is an effective therapeutic means for the treatment of these malignancies.[76] For example, microRNAs have been used to decrease the translation of crk and effectively inhibit the invasion in non-small cell lung carcinoma cell lines.[81]

$p15^{PAF}$ (pE-DB ID 6AAA) is a proliferating cell nuclear antigen (PCNA) associated factor.[82] It is localized primarily in the nucleus and shares the conserved PCNA binding motif with several other PCNA binding proteins including CDK inhibitor p21. It also binds the transactivation region of p53 and strongly inhibits its transcriptional activity. The expression of $p15^{PAF}$ in several types of tumor tissues was notably increased, especially in esophageal tumors. The structural characterization of human $p15^{PAF}$ showed that it is an IDP with nonrandom structural preferences at sites of interaction with other proteins,[79] suggesting $p15^{PAF}$ to be potential drug target.

$p27^{Kip1}$ (pE-DB ID 2AAA) is a human homologue of Sic1, both being pivotal CDK inhibitors and tight modulators of CDK-dependent phenotypes. $p27^{Kip1}$ mainly stops or slows down the cell division cycle, and thus plays an essential role in key cellular processes such as proliferation, differentiation and apoptosis.[83] If the expression of $p27^{Kip1}$ is reduced, the progression from G1 to S-phase becomes out of control, which facilitates the formation and growth of tumors. Therefore, $p27^{Kip1}$ is a tumor suppressor protein, and drugs able to protect/enhance the role of $p27^{Kip1}$ may be an effective means for anticancer strategies.[83] In this aspect, design of allosteric effectors (allosteric drugs) would be very useful, which has gaining a lot of momentum in traditional drug discovery.[84,85] On the other hand, the downregulation of $p27^{Kip1}$ aids maintenance of stem cell pluripotency and tissue regeneration.[86] For example, $p27^{Kip1}$ inhibition therapy has been proposed for hearing restoration in mammals. Recently, a high-throughput screening strategy has been applied to successfully identify novel $p27^{Kip1}$ transcriptional inhibitors.[86]

Optimism versus obstacles for drug design on IDPs

The results obtained in this study suggest that the druggability of IDPs may be comparable with that of ordered proteins. The average probability for cavities to be predicted druggable is 9% in IDPs, almost double the value found for ordered proteins (5%). Taking into consideration the high content of IDPs in various proteomes and their essential role in human diseases, we are optimistic on the design of drugs that target IDPs. Despite being in its infancy, the drug design against IDPs is in a continuous progress and essential advance has been achieved in a few cases. It is expected that the study will be extended and have a great future.

On the other hand, there are some obstacles for drug design targeting IDPs. The major obstacle is the lack of well-developed strategies. Traditional rational drug design against ordered proteins relies on the knowledge of the three-dimensional protein structure. However, IDPs usually exist in highly dynamic conformational ensembles, and accurate ensembles are

difficult to determine via either experimental or theoretical means, which exclude traditional approaches in most cases. As proof, most cases of drug designs for IDPs were carried out by experimental screening, but not via rational design. An additional obstacle is the specificity/promiscuity.[32,36,87,88] For IDPs with a determined conformational ensemble, a straightforward strategy of rational design is to extract metastable structures and then conduct traditional approaches, but the promiscuity would be serious because ligands bind to IDPs in a way of "ligand clouds around protein clouds."[18] Therefore, the development of novel strategies is needed for better rational drug design on IDPs.

Materials and Methods

Datasets

To examine the cavity properties of IDPs and compare them with those of ordered proteins, we constructed/adopted three datasets: Disprot-pdb, pE-DB and CavityTEST.

Disprot-pdb was constructed by combining information from the Database of Protein Disorder (DisProt)[89] and the Protein Data Bank (PDB).[90] We checked all records in DisProt with PDB links. The disorder ratios of the proteins in DisProt lie between 0 and 100%, and the structures solved in the PDB may reside in either ordered or disordered regions of the proteins. Therefore, we constructed our Disprot-pdb dataset by selecting proteins with at least 50% of solved residues in the PDB structure being labeled disordered in DisProt. The Leukemia fusion target AF9 (with PDB ID 2LM0) was also included into Disprot-pdb, which is an intrinsically disordered transcriptional regulator.[63] The dataset contains 37 proteins, and are listed in Table I. Among them, 19 are in monomer free state, 13 come from heterocomplexes, and 5 from homocomplexes.

pE-DB was adopted from Varadi *et al*., which provided structural ensembles of IDPs based on nuclear magnetic resonance (NMR) spectroscopy, small-angle X-ray scattering (SAXS), and other data measured in solution.[30] Ensembles in pE-DB usually consist of a few dozen to hundreds (and possibly even more) of conformers. We also included the oncoprotein c-Myc into the dataset for analysis. C-Myc is one of a few examples of IDP drug design that are widely discussed in the literature (see the Discussion section for details). Recently, large-scale molecular dynamics simulations have been conducted to determine the conformation ensemble of c-Myc$_{370–409}$, which showed agreement with experimental NMR data.[18] We incorporated them into our analysis. In total, the pE-DB dataset we used contained 16 entries, and are given in Table II.

CavityTEST is a test set used by the binding cavity detection program CAVITY,[25] which was originally collected by Q-SiteFinder developers.[31] CavityTEST contains 35 structurally distinct ordered proteins in the unbound state that share structural similarity with 35 proteins in the ligand-bound dataset which were determined to bind ligands. The dataset is listed in Table III.

Cavity calculations

The detection of binding cavities and the determination of their properties were conducted using the program CAVITY developed by Yuan *et al*.[25] Here we provide a very brief introduction on CAVITY. CAVITY used a probe sphere (with a default radius value 10 Å) to roll around the protein surface to detect the inaccessible volume (cavities). Since the cavities detected in this way are typically connected by the shallow grooves distributed around the rough surface, CAVITY adopted a shrink-and-expand algorithm to remove the linkage area within a depth threshold to separate the cavities, where the depth is defined as the distance from the surface to the bottom of cavities. A minimal depth parameter (8 Å as default) was also used to eliminate any cavities with too small depth, and a maximal joint depth parameter (20 Å as default) was used to restrain cavities with too large depth. For each of the resulting cavities, a quantity termed CavityScore was calculated based on their geometrical structure and physical chemistry properties,[25] for example, cavity volume, hydrophobic volume, cavity surface area, and hydrogen-bond forming surface area. The parameters of CAVITY were optimized by training on a refined set containing 1300 protein-ligand complexes from the PDBBind Database.[91] A binding site prediction test were performed[25] using 134 structures prepared by Q-SiteFinder developers,[31] where the success rate of CAVITY was 86% when only the first-ranked predicted cavities were considered, and the success rate increased to 96% if considering the true binding sites are among the top three predicted binding cavities. Such success rates of CAVITY are higher than other popular binding sites detection approaches, for example, LIGSITEcsc,[92] Q-sitefinder,[31] SURFNET,[93] and PASS.[94] Plotting the binding affinity values of 210 complexes versus the calculated CavityScore values of their binding cavities revealed a rough linear relationship between them, based on which an average pK_d can be predicted for any detected cavity,[25] which can be understood as the expected pK_d of the detected cavity with "properly designed ligands." For more details, refer to the original paper of Yuan *et al*.[25]

In our study, conformations from a PDB file were separated and were fed to CAVITY following the removal of water molecules. For complexes in Disprot-pdb formed by IDPs and ordered proteins, the ordered proteins were discarded. Default parameters of CAVITY were used in the calculations. For

N_{conf} conformations of a protein, a weight of $1/N_{conf}$ was assigned to each conformation in the property statistics so that the calculated average properties of datasets will not be dominated by proteins whose available conformation number is large. Residues were considered to belong to a single segment if they were involved in constituting the same cavity and their positions in the polypeptide chain were continuous.

The similarity between two cavities i and j is measured by the proportion of common atoms (p_{com}) among them and the root mean square deviation (RMSD) of these common atoms. To measure the conservation of a particular cavity i in an ensemble with distinct conformations, that is, whether the cavity i is present in many conformations, we picked out a cavity from each conformation J which possesses the highest p_{com} with i among all cavities of the conformation J, and calculated the average p_{com} and RMSD values of these picked cavities with the cavity i. The larger average p_{com} and smaller average RMSD, the more conservative the cavity is. The proteins with only one conformation in the datasets were excluded from the conservation analysis.

Conclusion

In this study, we conducted analyses on the binding cavities and druggability of IDPs, and compared them with those of ordered proteins. IDPs were shown to possess more binding cavities than ordered proteins of similar length, and the distribution of the cavity number caused by the conformation ensemble is very wide for IDPs. The cavity geometries for IDPs and ordered proteins also differ, for example, the cavities of IDPs have larger surface areas and volumes on average, and are more likely to be composed of a single segment. Most importantly, the predicted druggability of the cavities of IDPs is comparable with that of ordered proteins. Last, we have briefly discussed successful drug design examples, the potent drug targets, and the optimism *vs*. obstacles for drug design targeting IDPs.

Acknowledgments

The authors thank Prof. Luhua Lai, Dr. Yongqi Huang, and Chen Yu for helpful discussions.

References

1. Uversky VN (2013) A decade and a half of protein intrinsic disorder: biology still waits for physics. Protein Sci 22:693–724.
2. Huang YQ, Liu ZR (2010) Intrinsically disordered proteins: the new sequence-structure-function relations. Acta Phys Chim Sin 26:2061–2072.
3. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CR, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang CH, Kissinger CR, Bailey RW, Griswold MD, Chiu M, Garner EC, Obradovic Z (2001) Intrinsically disordered protein. J Mol Graph Model 19:26–59.
4. Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol 293:321–331.
5. Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol 323:573–584.
6. Cheng Y, LeGall T, Oldfield CJ, Mueller JP, Van YYJ, Romero P, Cortese MS, Uversky VN, Dunker AK (2006) Rational drug design via intrinsically disordered protein. Trends Biotechnol 24:435–442.
7. Metallo SJ (2010) Intrinsically disordered proteins are potential drug targets. Curr Opin Chem Biol 14:481–488.
8. Xie HB, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Uversky VN, Obradovic Z (2007) Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. J Proteome Res 6:1882–1898.
9. Midic U, Oldfield CJ, Dunker AK, Obradovic Z, Uversky VN (2009) Protein disorder in the human diseasome: unfoldomics of human genetic diseases. BMC Genomics 10:12.
10. Babu MM, van der Lee R, de Groot NS, Gsponer J (2011) Intrinsically disordered proteins: regulation and disease. Curr Opin Struct Biol 21:432–440.
11. Mendoza-Espinosa P, Garcia-Gonzalez V, Moreno A, Castillo R, Mas-Oliva J (2009) Disorder-to-order conformational transitions in protein structure and its relationship to disease. Mol Cell Biochem 330:105–120.
12. Cheng YG, LeGall T, Oldfield CJ, Dunker AK, Uversky VN (2006) Abundance of intrinsic disorder in protein associated with cardiovascular disease. Biochemistry 45:10448–10460.
13. Uversky VN, Oldfield CJ, Dunker AK (2008) Intrinsically disordered proteins in human diseases: introducing the D^2 concept. Annu Rev Biophys 37:215–246.
14. Dunker AK, Uversky VN (2010) Drugs for 'protein clouds': targeting intrinsically disordered transcription factors. Curr Opin Pharmacol 10:782–788.
15. Wang JH, Cao ZX, Zhao LL, Li SQ (2011) Novel strategies for drug discovery based on intrinsically disordered proteins (IDPs). Int J Mol Sci 12:3205–3219.
16. Chen CYC, Tou WL (2013) How to design a drug for the disordered proteins. Drug Discov Today 18:910–915.
17. Yuan YX, Pei JF, Lai LH (2011) LigBuilder 2: a practical de novo drug design approach. J Chem Inf Model 51:1083–1091.
18. Jin F, Yu C, Lai LH, Liu ZR (2013) Ligand clouds around protein clouds: a scenario of ligand binding with intrinsically disordered proteins. PLoS Comput Biol 9:e1003249.
19. Chene P (2004) Inhibition of the p53-MDM2 interaction: targeting a protein-protein interface. Mol Cancer Res 2:20–28.
20. Erkizan HV, Kong YL, Merchant M, Schlottmann S, Barber-Rotenberg JS, Yuan LS, Abaan OD, Chou TH, Dakshanamurthy S, Brown ML, Uren A, Toretsky JA (2009) A small molecule blocking oncogenic protein EWS-FLI1 interaction with RNA helicase A inhibits growth of Ewing's sarcoma. Nat Med 15:750–757.
21. Hammoudeh DI, Follis AV, Prochownik EV, Metallo SJ (2009) Multiple independent binding sites for small-molecule inhibitors on the oncoprotein c-Myc. J Am Chem Soc 131:7390–7401.

22. Srinivasan RS, Nesbit JB, Marrero L, Erfurth F, LaRussa VF, Hemenway CS (2004) The synthetic peptide PFWT disrupts AF4-AF9 protein complexes and induces apoptosis in t(4;11) leukemia cells. Leukemia 18:1364–1372.
23. Sievers SA, Karanicolas J, Chang HW, Zhao A, Jiang L, Zirafi O, Stevens JT, Munch J, Baker D, Eisenberg D (2011) Structure-based design of non-natural amino-acid inhibitors of amyloid fibril formation. Nature 475:96–100.
24. Yuzwa SA, Macauley MS, Heinonen JE, Shan XY, Dennis RJ, He YA, Whitworth GE, Stubbs KA, McEachern EJ, Davies GJ, Vocadlo DJ (2008) A potent mechanism-inspired O-GlcNAcase inhibitor that blocks phosphorylation of tau in vivo. Nat Chem Biol 4:483–490.
25. Yuan YX, Pei JF, Lai LH (2013) Binding site detection and druggability prediction of protein targets for structure-based drug design. Curr Pharm Des 19: 2326–2333.
26. Huang YQ, Liu ZR (2009) Kinetic advantage of intrinsically disordered proteins in coupled folding-binding process: a critical assessment of the "fly-casting" mechanism. J Mol Biol 393:1143–1159.
27. Gunasekaran K, Tsai CJ, Kumar S, Zanuy D, Nussinov R (2003) Extended disordered proteins: targeting function with less scaffold. Trends Biochem Sci 28:81–85.
28. Meszaros B, Tompa P, Simon I, Dosztanyi Z (2007) Molecular principles of the interactions of disordered proteins. J Mol Biol 372:549–561.
29. Huang YQ, Liu ZR (2010) Smoothing molecular interactions: the "kinetic buffer" effect of intrinsically disordered proteins. Proteins 78:3251–3259.
30. Varadi M, Kosol S, Lebrun P, Valentini E, Blackledge M, Dunker AK, Felli IC, Forman-Kay JD, Kriwacki RW, Pierattelli R, Sussman J, Svergun DI, Uversky VN, Vendruscolo M, Wishart D, Wright PE, Tompa P (2014) pE-DB: a database of structural ensembles of intrinsically disordered and of unfolded proteins. Nucleic Acids Res 42:D326–D335.
31. Laurie ATR, Jackson RM (2005) Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. Bioinformatics 21:1908–1916.
32. Liu ZR, Huang YQ (2014) Advantages of proteins being disordered. Protein Sci 23:539–550.
33. Knott M, Best RB (2012) A preformed binding interface in the unbound ensemble of an intrinsically disordered protein: evidence from molecular simulations. PLoS Comput Biol 8:e1002605.
34. Uversky VN (2013) Unusual biophysics of intrinsically disordered proteins. Biochim Biophys Acta 1834:932–951.
35. Song JH, Ng SC, Tompa P, Lee KAW, Chan HS (2013) Polycation-pi interactions are a driving force for molecular recognition by an intrinsically disordered oncoprotein family. PLoS Comput Biol 9:e1003239.
36. Huang YQ, Liu ZR (2013) Do intrinsically disordered proteins possess high specificity in protein-protein interactions? Chem-Eur J 19:4462–4467.
37. Fuxreiter M, Simon I, Friedrich P, Tompa P (2004) Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. J Mol Biol 338:1015–1026.
38. Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker AK, Uversky VN (2006) Analysis of molecular recognition features (MoRFs). J Mol Biol 362: 1043–1059.
39. Disfani FM, Hsu WL, Mizianty MJ, Oldfield CJ, Xue B, Dunker AK, Uversky VN, Kurgan L (2012) MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. Bioinformatics 28:I75–I83.
40. Vassilev LT, Vu BT, Graves B, Carvajal D, Podlaski F, Filipovic Z, Kong N, Kammlott U, Lukacs C, Klein C, Fotouhi N, Liu EA (2004) In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. Science 303:844–848.
41. Yu X, Narayanan S, Vazquez A, Carpizo DR (2014) Small molecule compounds targeting the p53 pathway: are we finally making progress? Apoptosis 19:1055–1068.
42. Tovar C, Rosinski J, Filipovic Z, Higgins B, Kolinsky K, Hilton H, Zhao XL, Vu BT, Qing WG, Packman K, Myklebost O, Heimbrook DC, Vassilev LT (2006) Small-molecule MDM2 antagonists reveal aberrant p53 signaling in cancer: implications for therapy. Proc Natl Acad Sci USA 103:1888–1893.
43. Berg T, Cohen SB, Desharnais J, Sonderegger C, Maslyar DJ, Goldberg J, Boger DL, Vogt PK (2002) Small-molecule antagonists of Myc/Max dimerization inhibit Myc-induced transformation of chicken embryo fibroblasts. Proc Natl Acad Sci USA 99:3830–3835.
44. Shi J, Stover JS, Whitby LR, Vogt PK, Boger DL (2009) Small molecule inhibitors of Myc/Max dimerization and Myc-induced cell transformation. Bioorg Med Chem Lett 19:6038–6041.
45. Yin XY, Giap C, Lazo JS, Prochownik EV (2003) Low molecular weight inhibitors of Myc-Max interaction and function. Oncogene 22:6151–6159.
46. Zirath H, Frenzel A, Oliynyk G, Segerstrom L, Westermark UK, Larsson K, Persson MM, Hultenby K, Lehtio J, Einvik C, Pahlman S, Kogner P, Jakobsson PJ, Henriksson MA (2013) MYC inhibition induces metabolic changes leading to accumulation of lipid droplets in tumor cells. Proc Natl Acad Sci USA 110: 10258–10263.
47. Fletcher S, Prochownik EV (in press) Small-molecule inhibitors of the Myc oncoprotein. Biochim Biophys Acta.
48. Hong SH, Youbi SE, Hong SP, Kallakury B, Monroe P, Erkizan HV, Barber-Rotenberg JS, Houghton P, Uren A, Toretsky JA (2014) Pharmacokinetic modeling optimizes inhibition of the 'undruggable' EWS-FLI1 transcription factor in Ewing Sarcoma. Oncotarget 5:338–350.
49. Palermo CM, Bennett CA, Winters AC, Hemenway CS (2008) The AF4-mimetic peptide, PFWT, induces necrotic cell death in MV4–11 leukemia cells. Leuk Res 32:633–642.
50. Bennett CA, Winters AC, Barretto NN, Hemenway CS (2009) Molecular targeting of MLL-rearranged leukemia cell lines with the synthetic peptide PFWT synergistically enhances the cytotoxic effect of established chemotherapeutic agents. Leuk Res 33:937–947.
51. Watson VG, Drake KM, Peng Y, Napper AD (2013) Development of a high-throughput screening-compatible assay for the discovery of inhibitors of the AF4-AF9 interaction using AalphaScreen technology. Assay Drug Dev Technol 11:253–268.
52. Krishnan N, Koveal D, Miller DH, Xue B, Akshinthala SD, Kragelj J, Jensen MR, Gauss CM, Page R, Blackledge M, Muthuswamy SK, Peti W, Tonks NK (2014) Targeting the disordered C terminus of PTP1B with an allosteric inhibitor. Nat Chem Biol 10:558–566.
53. Fokkens M, Schrader T, Klarner FG (2005) A molecular tweezer for lysine and arginine. J Am Chem Soc 127:14415–14421.
54. Prabhudesai S, Sinha S, Attar A, Kotagiri A, Fitzmaurice AG, Lakshmanan R, Ivanova MI, Loo JA,

Klarner FG, Schrader T, Stahl M, Bitan G, Bronstein JM (2012) A novel "molecular tweezer" inhibitor of alpha-Synuclein neurotoxicity in vitro and in vivo. Neurotherapeutics 9:464–476.
55. Sinha S, Lopes DHJ, Du ZM, Pang ES, Shanmugam A, Lomakin A, Talbiersky P, Tennstaedt A, McDaniel K, Bakshi R, Kuo PY, Ehrmann M, Benedek GB, Loo JA, Klarner FG, Schrader T, Wang CY, Bitan G (2011) Lysine-specific molecular tweezers are broad-spectrum inhibitors of assembly and toxicity of amyloid proteins. J Am Chem Soc 133:16958–16969.
56. Toth G, Gardai SJ, Zago W, Bertoncini CW, Cremades N, Roy SL, Tambe MA, Rochet JC, Galvagnion C, Skibinski G, Finkbeiner S, Bova M, Regnstrom K, Chiou SS, Johnston J, Callaway K, Anderson JP, Jobling MF, Buell AK, Yednock TA, Knowles TPJ, Vendruscolo M, Christodoulou J, Dobson CM, Schenk D, McConlogue L (2014) Targeting the intrinsically disordered structural ensemble of alpha-Synuclein by small molecules as a potential therapeutic strategy for Parkinson's disease. PLoS One 9:87133.
57. Tompa P (2011) Unstructural biology coming of age. Curr Opin Struct Biol 21:419–425.
58. Cuchillo R, Michel J (2012) Mechanisms of small-molecule binding to intrinsically disordered proteins. Biochem Soc Trans 40:1004–1008.
59. Uversky VN (2012) Intrinsically disordered proteins and novel strategies for drug discovery. Expert Opin Drug Discov 7:475–488.
60. Joerger AC, Fersht AR (2008) Structural biology of the tumor suppressor p53. Annu Rev Biochem 77:557–582.
61. Li ZY, Ni M, Li JK, Zhang YP, Ouyang Q, Tang C (2011) Decision making of the p53 network: Death by integration. J Theor Biol 271:205–211.
62. Huang YQ, Liu ZR (2011) Anchoring intrinsically disordered proteins to multiple targets: Lessons from N-terminus of the p53 protein. Int J Mol Sci 12:1410–1430.
63. Leach BI, Kuntimaddi A, Schmidt CR, Cierpicki T, Johnson SA, Bushweller JH (2013) Leukemia fusion target AF9 is an intrinsically disordered transcriptional regulator that recruits multiple partners via coupled folding and binding. Structure 21:176–183.
64. Hegyi H, Buday L, Tompa P (2009) Intrinsic structural disorder confers cellular viability on oncogenic fusion proteins. PLoS Comput Biol 5:1000552.
65. Johnson TO, Ermolieff J, Jirousek MR (2002) Protein tyrosine phosphatase 1B inhibitors for diabetes. Nat Rev Drug Discov 1:696–709.
66. Tompa P (2014) Multisteric regulation by structural disorder in modular signaling proteins: an extension of the concept of allostery. Chem Rev 114:6715–6732.
67. Jin F, Liu ZR (2013) Inherent relationships among different biophysical prediction methods for intrinsically disordered proteins. Biophys J 104:488–495.
68. Zhu M, De Simone A, Schenk D, Toth G, Dobson CM, Vendruscolo M (2013) Identification of small-molecule binding pockets in the soluble monomeric form of the A beta 42 peptide. J Chem Phys 139:035101.
69. Hsu SY, Kaipia A, Zhu L, Hsueh AJW (1997) Interference of BAD (Bcl-xL/Bcl-2-associated death promoter)-induced apoptosis in mammalian cells by 14-3-3 isoforms and P11. Mol Endocrinol 11:1858–1867.
70. Li Z, Zhao B, Wang P, Chen F, Dong ZH, Yang HR, Guan KL, Xu YH (2010) Structural insights into the YAP and TEAD complex. Genes Dev 24:235–240.
71. Wang SR, Trumble WR, Liao H, Wesson CR, Dunker AK, Kang CH (1998) Crystal structure of calsequestrin from rabbit skeletal muscle sarcoplasmic reticulum. Nat Struct Biol 5:476–483.
72. Jin L, Briggs SL, Chandrasekhar S, Chirgadze NY, Clawson DK, Schevitz RW, Smiley DL, Tashjian AH, Zhang FM (2000) Crystal structure of human parathyroid hormone 1–34 at 0.9-angstrom resolution. J Biol Chem 275:27238–27244.
73. DiGiammarino EL, Filippov I, Weber JD, Bothner B, Kriwacki RW (2001) Solution structure of the p53 regulatory domain of the p19(Arf) tumor suppressor protein. Biochemistry 40:2379–2386.
74. Hayes PL, Lytle BL, Volkman BF, Peterson FC (2008) The solution structure of ZNF593 from Homo sapiens reveals a zinc finger in a predominately unstructured protein. Protein Sci 17:571–576.
75. Furst J, Schedlbauer A, Gandini R, Garavaglia ML, Saino S, Gschwentner M, Sarg B, Lindner H, Jakab M, Ritter M, Bazzini C, Botta G, Meyer G, Kontaxis G, Tilly BC, Konrat R, Paulmichl M (2005) ICln(159) folds into a pleckstrin homology domain-like structure. J Biol Chem 280:31276–31282.
76. Kobashigawa Y, Sakai M, Naito M, Yokochi M, Kumeta H, Makino Y, Ogura K, Tanaka S, Inagaki F (2007) Structural basis for the transforming activity of human cancer-related signaling adaptor protein CRK. Nat Struct Mol Biol 14:503–510.
77. Ellena JF, Liang BY, Wiktor M, Stein A, Cafiso DS, Jahn R, Tamm LK (2009) Dynamic structure of lipid-bound synaptobrevin suggests a nucleation-propagation mechanism for trans-SNARE complex formation. Proc Natl Acad Sci USA 106:20306–20311.
78. Sivakolundu SG, Bashford D, Kriwacki RW (2005) Disordered p27(Kip1) exhibits intrinsic structure resembling the Cdk2/cyclin A-bound conformation. J Mol Biol 353:1118–1128.
79. De Biasio A, de Opakua AI, Cordeiro TN, Villate M, Merino N, Sibille N, Lelli M, Diercks T, Bernado P, Blanco FJ (2014) p15(PAF) is an intrinsically disordered protein with nonrandom structural preferences at sites of interaction with other proteins. Biophys J 106:865–874.
80. Lee VMY, Goedert M, Trojanowski JQ (2001) Neurodegenerative tauopathies. Annu Rev Neurosci 24:1121–1159.
81. Crawford M, Brawner E, Batte K, Yu L, Hunter MG, Otterson GA, Nuovo G, Marsh CB, Nana-Sinkam SP (2008) MicroRNA-126 inhibits invasion in non-small cell lung carcinoma cell lines. Biochem Biophys Res Commun 373:607–612.
82. Yu PW, Huang B, Shen M, Lau C, Chan E, Michel J, Xiong Y, Payan DG, Luo Y (2001) p15(PAF), a novel PCNA associated factor with increased expression in tumor tissues. Oncogene 20:484–489.
83. Borriello A, Bencivenga D, Criscuolo M, Caldarelli I, Cucciolla V, Tramontano A, Borgia A, Spina A, Oliva A, Naviglio S, Della Ragione F (2011) Targeting p27(Kip1) protein: its relevance in the therapy of human cancer. Expert Opin Ther Targets 15:677–693.
84. Wang Q, Qi YF, Yin N, Lai LH (2014) Discovery of novel allosteric effectors based on the predicted allosteric sites for *Escherichia coli* D-3-phosphoglycerate dehydrogenase. PLoS One 9:e94829.
85. Herbert C, Schieborr U, Saxena K, Juraszek J, De Smet F, Alcouffe C, Bianciotto M, Saladino G, Sibrac D, Kudlinzki D, Sreeramulu S, Brown A, Rigon P, Herault JP, Lassalle G, Blundell TL, Rousseau F, Gils A, Schymkowitz J, Tompa P, Herbert JM, Carmeliet P, Gervasio FL, Schwalbe H, Bono F (2013) Molecular mechanism of SSR128129E, an extracellularly acting, small-molecule, allosteric inhibitor of FGF receptor signaling. Cancer Cell 23:489–501.
86. Walters BJ, Lin WW, Diao SY, Brimble M, Iconaru LI, Dearman J, Goktug A, Chen TS, Zuo J (2014) High-

throughput screening reveals alsterpaullone, 2-cyanoethyl as a potent p27(Kip1) transcriptional inhibitor. PLoS One 9:e91173.
87. Bhattacherjee A, Wallin S (2013) Exploring protein-peptide binding specificity through computational peptide screening. PLoS Comput Biol 9:e1003277.
88. Zhou HX (2012) Intrinsic disorder: signaling via highly specific but short-lived association. Trends Biochem Sci 37:43–48.
89. Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, Szabo B, Tompa P, Chen J, Uversky VN, Obradovic Z, Dunker AK (2007) DisProt: the database of disordered proteins. Nucleic Acids Res 35:D786–D793.
90. Berman H, Henrick K, Nakamura H, Markley JL (2007) The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. Nucleic Acids Res 35:D301–D303.
91. Wang RX, Fang XL, Lu YP, Wang SM (2004) The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures. J Med Chem 47:2977–2980.
92. Hendlich M, Rippmann F, Barnickel G (1997) LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. J Mol Graph Model 15:359–363.
93. Laskowski RA (1995) SURFNET - a program for visualizing molecular-surfaces, cavities, and intermolecular interactions. J Mol Graph 13:323–330.
94. Brady GP, Stouten PFW (2000) Fast prediction and visualization of protein binding pockets with PASS. J Comput-Aided Mol Des 14:383–401.